



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

ON A SINGLE NUMERICAL INDEX OF THE AGE DISTRIBUTION
OF A POPULATION¹

BY RAYMOND PEARL

DEPARTMENT OF BIOMETRY AND VITAL STATISTICS, JOHNS HOPKINS UNIVERSITY

Read before the Academy, April 27, 1920

It is an obvious fact that the crude death rate of any community is influenced in a marked degree by the age distribution of the living population of that community. Before any critical conclusions about the true force of mortality can be drawn some sort of correction must be applied to take account of the age distribution of the population. If one desires to make any analysis of the correlation of death rates with each other or with sundry environmental factors it is an absolutely essential prerequisite that there be found some single numerical expression which will be an approximate index of the age distribution of the population of each locality dealt with.

From a mathematical viewpoint the problem presented is one incapable of *exact* solution. The problem mathematically is to find a single parameter which will fully describe so complex a curve as that of the age distribution of a population, the general form of which is indicated in figure 1. But this is impossible. No single parameter can possibly describe fully such a curve. The best that can be done is to approximate as closely as may be to the impossible ideal. It is the purpose of this paper to describe and illustrate an approximation which comes very close to the requirements, indeed quite sufficiently so for all practical statistical purposes in all cases where it has yet been tested out, and is easy to calculate.

In a recent paper,² I proposed as an index of differences in age composition of populations, and applied it to 40 American cities, the expression

$$\chi^2 = S\left(\frac{\Delta^2}{P}\right) \quad (1)$$

"where Δ is the deviation for each of six age groups (viz, 0-4, 5-14, 15-24, 25-44, 45-64, 65 and over) of the percentage of the actual population of each city in 1910 in each age group, from the percentage in the same group in the Standard Population of Glover's³ Life Table, denoted in the formula by P . S denotes summation of all six values. The value χ^2 measures the extent to which each city deviates in the age constitution of its population from a fixed standard, but does not tell the nature or kind of the deviation."

In the paper cited it was shown empirically that in spite of the obvious defect of this index mentioned in the last sentence, it was adequate for the restricted material and purposes there involved. Since that paper was published it has occurred to me that the outstanding defect of the former index will, to that degree of accuracy which is requisite for all practical

statistical purposes, be entirely removed by adopting as an age-constitution index the function

$$\phi = S \left\{ \frac{\Delta^2}{P} \right\} (M - M_p) \quad (2)$$

where S , Δ and P have the same significance as before, and M = mean age of living population in any community, M_p = mean age of persons in a stationary population unaffected by migration and which, assuming

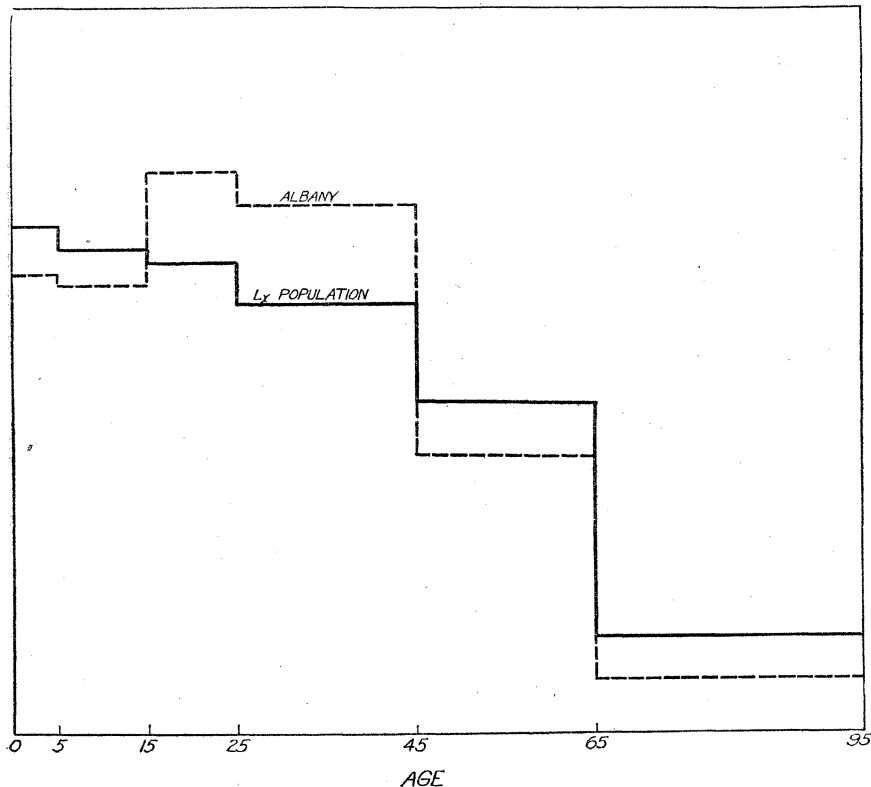


FIG. 1. Diagram showing age distribution of population, grouped in six age classes.

The last class in this and other diagrams is assumed to have its ending at age 95.

The solid line gives the stationary life table population, and the broken line the population of Albany. Total area under each curve = 100%.

the mortality rates of Glover's Life Table, would result if 100,000 persons were born alive uniformly throughout each year (M_p calculated from L_x line of Glover's Table (p. 16) = 33.796 years).

This procedure simply multiplies our former index χ^2 by the difference (given its proper sign) between the mean age of the observed population and the mean age of the standard population on the basis of which χ^2 was calculated. Since in fact the mean age of any actual urban population is never likely to be as great as the mean age of the stationary popula-

tion chosen as a standard of reference the actual values of ϕ will practically always be negative for cities. The smaller these negative values are numerically the greater will be the proportion of older persons in the population concerned. In short this function ϕ tells us not only the degree to which a given population deviates in its age distribution from a fixed standard age distribution, but also the nature of this deviation, whether on the one hand in the direction of a relative excess of aged, or

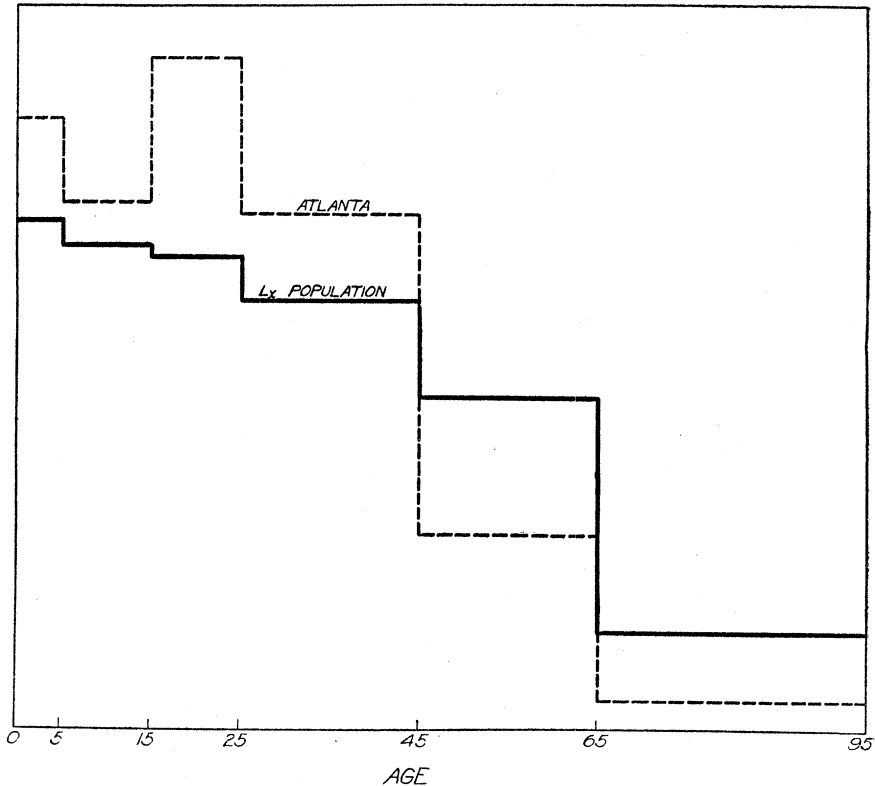


FIG. 2. Diagram showing age distribution of population, grouped in six age classes. The solid line gives the stationary life-table population, and the broken line the population of Atlanta. Total area under each curve = 100%.

on the other hand in the direction of a relative excess of the young. Theoretically it is possible for two populations differing from one another in a compensatory way to give the same values for the index ϕ . But two populations which differ in age distribution in any fundamental respect which could affect appreciably crude death rates will, in all populations I have been able to test, give different values of ϕ , provided the age classification from which the function is calculated is finely enough divided.

II

The numerical values of the index ϕ for a series of American cities, on the basis of the age distribution of the 1910 census are given by way of illustration in table 1. The cities are arranged in descending order of departure from the L_x population.

The extent of the departures from the standard population denoted by the several values of the indices may be indicated graphically by plotting the age distribution curve, on a percentage basis, for the two extreme cities in the table, Albany and Atlanta, against the L_x population. This is done in figures 1 and 2.

The general characteristics of an urban population as compared with a stationary life table population are well shown in these diagrams. In general a city has population under 40 to 45 in excess, and population older than that in defect. The reason is obvious.

TABLE 1
SHOWING AGE-CONSTITUTION INDICES OF THE POPULATION OF AMERICAN CITIES

CITY	ϕ	CITY	ϕ
Albany.....	- 10.73	Lowell.....	- 34.91
Los Angeles.....	- 13.61	Philadelphia.....	- 39.51
Oakland.....	- 17.18	San Francisco.....	- 43.35
Washington.....	- 20.67	St. Louis.....	- 44.57
Cincinnati.....	- 22.93	Nashville.....	- 48.32
Dayton.....	- 24.10	Buffalo.....	- 48.61
Indianapolis.....	- 24.23	New Orleans.....	- 52.44
Rochester.....	- 26.62	Minneapolis.....	- 55.39
Grand Rapids.....	- 27.95	Milwaukee.....	- 62.17
Columbus.....	- 29.08	Newark.....	- 62.54
Providence.....	- 30.05	Chicago.....	- 68.76
Cambridge.....	- 30.26	St. Paul.....	- 68.95
Boston.....	- 31.05	Pittsburgh.....	- 71.77
Baltimore.....	- 31.80	Fall River.....	- 73.39
Louisville.....	- 32.19	New York.....	- 74.42
Toledo.....	- 33.21	Cleveland.....	- 74.51
New Haven.....	- 33.70	Atlanta.....	- 82.71

III

It will be well to see in detail how the age distributions of some cities having nearly the same value for ϕ compare in respect of actual age distribution of their populations. Table 2 furnishes such comparisons for four pairs of cities, covering fairly the whole rank of values of ϕ shown in table 1.

From this table it is seen that as we pass from cities having a value of ϕ of about 24 to cities having ϕ equal to about 74 there is a steady change from populations having relatively many persons in the higher age groups

TABLE 2
PERCENTAGE AGE DISTRIBUTION OF THE POPULATION OF CERTAIN CITIES IN 1910

AGE CLASSES (YEARS)	DAYTON	INDIAN- APOLIS	PROVI- DENCE	CAM- BRIDGE	NASH- VILLE	BUFFALO	NEW YORK	CLEVE- LAND
	$\phi = -24.10$	$\phi = -24.23$	$\phi = -30.05$	$\phi = -30.26$	$\phi = -48.32$	$\phi = -48.61$	$\phi = -74.42$	$\phi = -74.51$
0-4	9.1%	8.0%	9.7%	10.3%	9.2%	10.0%	10.6%	11.1%
5-14	15.4	15.3	16.5	17.5	17.8	18.3	18.1	17.4
15-24	19.5	19.4	19.0	18.4	21.9	20.6	20.8	20.5
25-44	34.6	35.9	34.2	33.3	32.2	32.3	33.9	34.4
45-64	17.0	17.0	16.2	16.0	15.1	15.5	13.7	13.4
65 & over	4.4	4.3	4.2	4.4	3.8	3.4	2.8	3.0
Totals	100.0	99.9	99.8	99.9	100.0	100.1	99.9	99.8

to populations having relatively few persons in those groups. Furthermore, it is evident that in each of the four pairs of cities compared the agreement between the two cities having nearly identical values of ϕ is very close in respect of actual percentage distribution of the population. Dayton and Indianapolis were for all practical statistical purposes identical in the age distribution of their populations in 1910. So were Providence and Cambridge, Nashville and Buffalo, and New York and Cleveland. In each case the curve for one distribution of the pair winds in and out about the path set by the other. Of course, we should get finer differentiations brought out by the ϕ function if we used 15 or 20 age classes instead of the 6 here employed. But for the purposes of the investigation in which I am using these functions, and for purposes of illustration of method 6 classes are sufficient. A word of caution must, however, be emphasized here. The reliability of ϕ as an approximate index of differences in age distributions of population is greater as we pass in either direction towards the ends of its range of values. In the case of populations giving values of ϕ near the mean (say in the thirties for American cities) it may be necessary in order to get really differentiant values to calculate from a rather fine age grouping.

In general it is believed that the function of an age distribution of a population here proposed will give, in a single numerical expression, a substantially accurate indication of the essential nature of that age distribution, and will facilitate the differentiation and classification of populations in respect of this characteristic for statistical studies, particularly by the method of multiple correlation.

¹ Papers from the Department of Biometry and Vital Statistics, School of Hygiene and Public Health, Johns Hopkins University, No. 16.

² Pearl, R., "Influenza Studies. I—On Certain General Statistical Aspects of the 1918 Epidemic in American Cities," *Public Health Repts.*, **34**, 1919 (1743-1783).

³ Glover, J. W., *United States Life Tables*, 1910, Bureau of the Census, 1916.